A FAST VARIATIONAL BAYES ALGORITHM FOR SPARSE SEMI-SUPERVISED UNMIXING OF OMEGA/MARS EXPRESS DATA

Athanasios A. Rontogiannis[†]*, Konstantinos Themelis*[†]*, Olga Sykioti*[†]*, Konstantinos Koutroumbas*[†]

[†]IAASARS, National Observatory of Athens, 152 36, P. Penteli, Greece
{tronto,themelis,sykioti,koutroum}@noa.gr

ABSTRACT

In this paper a variational Bayesian framework for semi-supervised unmixing of hyperspectral data is proposed. It is assumed that a set of spectral signatures of materials *possibly* present in the hyperspectral image is given *a priori*. A hierarchical Bayesian model is adopted, incorporating prior distributions that both promote sparsity and satisfy the non-negativity constraints for the abundance coefficients. By exploiting the special form of the resulting posterior distributions a computationally efficient variational Bayes (VB) algorithm is derived to perform Bayesian inference. Experimental results conducted on hyperspectral data from the OMEGA sensor on board ESA's Mars Express satellite demonstrate both the performance and the sparsity-promoting nature of the proposed VB algorithm.

Index Terms— semi-supervised hyperspectral unmixing, linear sparse regression, Bayesian inference, planetology, OMEGA data

1. INTRODUCTION

Spectral unmixing of hyperspectral data has received considerable attention in the scientific literature during recent years, e.g. [1]. The linear mixing model is widely used to describe the interrelation among the disparate materials' spectra (*endmembers*) in a single pixel. Nevertheless, the prior knowledge of the exact endmembers present in each pixel is not always available. To complement this prior knowledge, spectral libraries have been utilized in the unmixing process, that contain the spectral information of hundreds of materials. This wealth of endmember availability calls for the development of *semi-supervised* unmixing techniques.

Semi-supervised unmixing is the process of determining how many and which endmembers are present in each pixel of a hyperspectral remote scene, as well as estimating their proportional contributions (*abundances*). A reasonable assumption is that only a few of the available endmembers will contribute in the spectrum of a single mixed pixel. Hence, semi-supervised unmixing is bound up with *sparsity* in the abundance vector. Many works have recently considered the unmixing process as a sparse linear regression problem, belonging either to the deterministic or the probabilistic framework. Characteristic examples of deterministic methods include [2, 3]. In [2], sparse unmixing is addressed via variable splitting and the augmented Lagrangian method of multipliers (SUnSAL). In [3], a variation of the former method is proposed, termed sparse unmixing via variable splitting augmented Lagrangian and total variation (SUnSAL-TV), which includes spatial information on the formulation of the sparse unmixing problem by means of a TV regularizer. In the probabilistic setting, a Bayesian treatment is provided in [4], where a suitable, sparsity promoting Laplace prior is utilized to model the abundance vector. Bayesian inference is then performed using a first-moments approximating VB algorithm.

In this paper, we propose a Bayesian treatment of the sparse linear unmixing problem. First, we adopt the hierarchical Bayesian model of [4], which is shown to promote sparsity and simultaneously meet the non-negativity constraint of the abundance vector (imposed due to physical restrictions). Next, we rely on a VB algorithm to perform statistical inference for the model parameters. Closed form expressions are provided for the updating of the parameters of all posterior approximating distributions. More importantly, based on suitable algebraic manipulations, a fast scheme is derived that allows us to reduce the computational complexity of the VB algorithm by one order of magnitude. This scheme performs Bayesian inference for all model parameters, and hence, there is no need for parameter cross-validation (as opposed to deterministic methods, e.g. SUnSAL, [2]). To demonstrate the efficiency of the proposed scheme, experimental results on a typical Martian hyperspectral image of the Syrtis Major volcanic complex are provided. This area is of great interest and has recently been utilized in the evaluation of sparse unmixing algorithms, [5].

2. PROBLEM FORMULATION

Each pixel of a hyperspectral image can be represented by a M-dimensional vector $\mathbf{y}$, whose elements are reflectance values measured in the corresponding spectral bands. The linear mixture model (LMM) adopted here assumes that the spectrum of a mixed pixel is a linear combination of component spectra, i.e.

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \mathbf{n} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_N]\,\mathbf{w} + \mathbf{n}, \tag{1}$$

where the $M \times 1$ dimensional vector $\boldsymbol{\phi}_i$ represents the spectral signature (i.e., the reflectance values in all spectral bands) of the ith endmember, $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T$ is the $N \times 1$ abundance vector associated with $\mathbf{y}$, and $\mathbf{n}$ is a N-dimensional vector of random zero mean Gaussian noise with covariance matrix $\beta^{-1}\mathbf{I}$. Due to physical restrictions, a non-negativity constraint is imposed on the abundance vector $\mathbf{w}$, i.e., $w_i \geq 0,\ i = 1, 2, \ldots, N$. Moreover, a natural assumption is that only a few of the N possible endmembers contribute in each pixel $\mathbf{y}$, hence, the abundance vector $\mathbf{w}$ is inherently *sparse*, i.e., $\|\mathbf{w}\|_0 = \xi \ll N$, where $\|\mathbf{w}\|_0$ is the number of nonzero components of $\mathbf{w}$.

The objective of semi-supervised unmixing is to estimate the abundance vector $\mathbf{w}$ given the pixel's measured spectra $\mathbf{y}$ and a set of possible endmember signatures $\mathbf{\Phi}$, subject to the positivity constraint and the sparsity assumption. To this end, a Bayesian approach

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: ARISTEIA- HSI-MARS-1413.

is presented in this paper, where the hierarchical Bayesian model of [4] is adopted and Bayesian inference is performed using a fast VB algorithm.

3. BAYESIAN MODELING

The Bayesian model explored in this paper is that proposed in [4]. Its key feature is the utilization of a conjugate-friendly expression of the double exponential (Laplace) distribution, which is widely used to promote sparsity. To begin with the description of the model, the presence of Gaussian noise in (1) dictates that the likelihood function of the data $\mathbf{y}$ is

$$\begin{aligned} p(\mathbf{y}|\mathbf{w},\beta) &= \mathcal{N}(\mathbf{y}|\mathbf{\Phi}\mathbf{w},\beta^{-1}\mathbf{I}_M) \\ &= (2\pi)^{-\frac{M}{2}}\,\beta^{\frac{M}{2}}\exp\left[-\frac{\beta}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{w}\|_2^2\right]. \end{aligned} \tag{2}$$

Due to space limitations, we briefly report the priors of the probabilistic model parameters defined in [4]. To satisfy the non-negativity constraint, the prior of the abundance vector $\mathbf{w}$ is expressed as

$$p(\mathbf{w}|\boldsymbol{\gamma},\beta) = \mathcal{N}_{\mathbf{R}_+^N}(\mathbf{w}|\mathbf{0},\beta^{-1}\mathbf{\Gamma}), \tag{3}$$

where $\mathbf{R}_+^N$ is the non-negative orthant of $\mathcal{R}^N$, $\mathcal{N}_{\mathbf{R}_+^N}(\cdot)$ stands for the N-variate normal distribution $\mathcal{N}(\cdot)$ truncated in $\mathbf{R}_+^N$ (see also [4]), and $\mathbf{\Gamma}$ is a $N\times N$ diagonal matrix with $\mathbf{\Gamma}=\mathrm{diag}(\gamma_1,\gamma_2,\ldots,\gamma_N)$. The variances $\gamma_i, i=1,2,\ldots,N$ are Gamma distributed, i.e.,

$$p(\gamma_i|\lambda_i) = \Gamma(\gamma_i|1,\frac{\lambda_i}{2}) = \frac{\lambda_i}{2}\exp\left[-\frac{\lambda_i}{2}\gamma_i\right], \tag{4}$$

where λ_i's are hyperparameters that control the level of sparsity. By combining the (3) and (4) the overall prior of $\mathbf{w}$ is a non-negatively truncated Laplace distribution, as shown in [4]. To infer the value of λ_i's also from the data, we further assign a Gamma distribution to them parameterized by $\rho,\delta\geq 0$, with $\rho,\delta\simeq 0$, i.e.,

$$p(\lambda_i|\rho,\delta) = \Gamma(\lambda_i|\rho,\delta) = \frac{\delta^\rho}{\Gamma(\rho)}{\lambda_i}^{\rho-1}\exp\left[-\delta\lambda_i\right]. \tag{5}$$

Finally, a Gamma prior distribution parameterized by $\kappa,\theta\geq 0$, with $\kappa,\theta\simeq 0$ is selected over the noise precision β, defined as

$$p(\beta|\kappa,\theta) = \Gamma(\beta|\kappa,\theta) = \frac{\theta^\kappa}{\Gamma(\kappa)}\beta^{\kappa-1}\exp\left[-\theta\beta\right], \tag{6}$$

where $\beta\geq 0$. The mean and variance of the Gamma distribution are $\mathrm{E}[p(\beta|\kappa,\theta)]=\frac{\kappa}{\theta}$, and $var[p(\beta|\kappa,\theta)]=\frac{\kappa}{\theta^2}$, respectively.

4. BAYESIAN INFERENCE

According to the Bayes' rule, the posterior distribution of $\mathbf{w},\beta,\boldsymbol{\gamma},\boldsymbol{\lambda}$ is expressed as

$$p\left(\mathbf{w},\beta,\boldsymbol{\gamma},\boldsymbol{\lambda}|\mathbf{y}\right) = \frac{p\left(\mathbf{y}|\mathbf{w},\beta\right)p\left(\mathbf{w}|\boldsymbol{\gamma},\beta\right)p\left(\boldsymbol{\gamma}|\boldsymbol{\lambda}\right)p\left(\boldsymbol{\lambda}\right)p\left(\beta\right)}{\int p\left(\mathbf{y},\mathbf{w},\beta,\boldsymbol{\gamma},\boldsymbol{\lambda}\right)d\mathbf{w}d\boldsymbol{\gamma}d\boldsymbol{\lambda}d\beta}. \tag{7}$$

However, the exact computation of the posterior is infeasible due to the integration at the denominator. In this paper we rely on the variational Bayesian framework [6] to approximate this joint posterior.

Assuming posterior independence among model parameters, the joint posterior (7) can be factorized as

$$p\left(\mathbf{w},\beta,\boldsymbol{\gamma},\boldsymbol{\lambda}|\mathbf{y}\right) \approx q\left(\mathbf{w},\beta,\boldsymbol{\gamma},\boldsymbol{\lambda}\right) = q(\mathbf{w})q(\beta)\prod_{i=1}^N q(\gamma_i)\prod_{i=1}^N q(\lambda_i), \tag{8}$$

and we can derive closed form expressions for all approximate posterior distributions $q(\mathbf{w}), q(\boldsymbol{\gamma}), q(\boldsymbol{\lambda})$, and $q(\beta)$, by utilizing the Kullback-Leibler (KL) distance minimization criterion, [6]. It is not difficult to verify by simple computations that the posterior $q(\mathbf{w})$ is a non-negatively truncated Gaussian distribution given by

$$q(\mathbf{w}) = \mathcal{N}_{\mathbf{R}_+^N}(\mathbf{w}|\boldsymbol{\mu},\mathbf{\Sigma}), \tag{9}$$

with

$$\boldsymbol{\mu} = \langle\beta\rangle\mathbf{\Sigma}\mathbf{\Phi}^T\mathbf{y}, \text{ and } \mathbf{\Sigma} = \langle\beta\rangle^{-1}\left(\mathbf{\Phi}^T\mathbf{\Phi}+\langle\mathbf{\Gamma}^{-1}\rangle\right)^{-1}, \tag{10}$$

where $\langle\cdot\rangle$ denotes expectation of a random variable with respect to its corresponding posterior $q(\cdot)$. The posterior $q(\beta)$ for the precision parameter β is expressed as

$$q(\beta) = \Gamma\left(\frac{M+N}{2}+\kappa, \frac{1}{2}\langle\|\mathbf{y}-\mathbf{\Phi}\mathbf{w}\|^2\rangle+\theta+\frac{1}{2}\langle\mathbf{w}^T\mathbf{\Gamma}^{-1}\mathbf{w}\rangle\right). \tag{11}$$

Straightforward computations yield that the approximating posterior pdf of $\gamma_i, i=1,2,\ldots,N$ is the following generalized inverse Gaussian distribution

$$\begin{aligned} q(\gamma_i) = &\left(\frac{\langle\lambda_i\rangle}{2\pi}\right)^{\frac{1}{2}} \\ &{\gamma_i}^{-\frac{1}{2}}\exp\left[-\frac{\langle\beta\rangle\langle w_i^2\rangle}{2\gamma_i}-\frac{\langle\lambda_i\rangle}{2}\gamma_i+\sqrt{\langle\beta\rangle\langle\lambda_i\rangle}\langle w_i\rangle\right]. \end{aligned} \tag{12}$$

Next, the posterior $q(\lambda_i), i=1,2,\ldots,N$ is expressed as

$$q(\lambda_i) = \Gamma\left(\alpha_i|1+\rho,\frac{\langle\gamma_i\rangle}{2}+\delta\right). \tag{13}$$

It is easy to verify from the resulting posterior distributions that the model parameters are interrelated. This gives rise to an iterative updating procedure, where the distributions' moments are easily evaluated using the following results

$$\langle w_i\rangle = \mu_{i,tr}, \qquad \langle w_i^2\rangle = \mu_{i,tr}^2+\sigma_{ii,tr} \tag{14}$$

$$\langle\gamma_i\rangle = \sqrt{\frac{\langle\beta\rangle\langle w_i^2\rangle}{\langle\lambda_i\rangle}}+\frac{1}{\langle\lambda_i\rangle}, \qquad \langle\gamma_i^{-1}\rangle = \sqrt{\frac{\langle\lambda_i\rangle}{\langle\beta\rangle\langle w_i^2\rangle}} \tag{15}$$

$$\langle\lambda_i\rangle = \frac{1+\rho}{\frac{1}{2}\langle\gamma_i\rangle+\delta} \tag{16}$$

$$\langle\|\mathbf{y}-\mathbf{\Phi}\mathbf{w}\|^2\rangle = \|\mathbf{y}-\mathbf{\Phi}\boldsymbol{\mu}_{tr}\|^2+\mathrm{Trace}\left[\mathbf{\Phi}\mathbf{\Sigma}_{\mathrm{tr}}\mathbf{\Phi}^{\mathrm{T}}\right] \tag{17}$$

$$\langle\mathbf{w}^T\mathbf{\Gamma}^{-1}\mathbf{w}\rangle = \sum_{i=1}^N\left[\langle\gamma_i^{-1}\rangle\langle w_i^2\rangle\right] \tag{18}$$

$$\langle\beta\rangle = \frac{\frac{M+N}{2}+\kappa}{\frac{1}{2}\langle\|\mathbf{y}-\mathbf{\Phi}\mathbf{w}\|^2\rangle+\theta+\frac{1}{2}\langle\mathbf{w}^T\mathbf{\Gamma}^{-1}\mathbf{w}\rangle}, \tag{19}$$

where $\boldsymbol{\mu}_{tr}=[\mu_{1,tr},\mu_{2,tr},\ldots,\mu_{N,tr}]^T$ is the mean and $\mathbf{\Sigma}_{tr}$ is the covariance matrix of the truncated Gaussian distribution $q(\mathbf{w})$ in (9).

Note that $\boldsymbol{\mu}_{tr}$ will be the estimate of the sparse abundance vector of the pixel $\mathbf{y}$. The proposed VB scheme iterates among the parameters of the approximating posterior distributions $q(\mathbf{w})$, $q(\gamma_i)$, $q(\lambda_i)$, $q(\beta)$, utilizing the required moments in (14)-(19). Convergence is achieved since in each step the KL distance between the true posterior (7) and the approximating distribution (8) is decreased. The most computationally demanding tasks of the proposed VB algorithm involve the computation of $\boldsymbol{\mu}_{tr}$ and $\boldsymbol{\Sigma}_{tr}$ of the truncated Gaussian distribution (9). To reduce complexity significantly, an efficient scheme is proposed next for the computation of $\boldsymbol{\mu}_{tr}$. Moreover, the need to compute $\boldsymbol{\Sigma}_{tr}$ analytically is alleviated by making the reasonable approximations $\langle \|\mathbf{y}-\boldsymbol{\Phi}\mathbf{w}\|^2 \rangle = \|\mathbf{y}-\boldsymbol{\Phi}\boldsymbol{\mu}_{tr}\|^2$ and $\langle w_i^2 \rangle = \mu_{i,tr}^2$.

4.1. Fast computation of the abundance vector estimate $\boldsymbol{\mu}_{tr}$

In [4], an iterative scheme has been proposed to compute the expectation of a multivariate Gaussian distribution truncated in the non-negative orthant of $\mathcal{R}^N$. In this paper, we propose a more computationally efficient implementation of this scheme, based on suitable algebraic manipulations. The scheme proposed in [4] iterates among the means of the one-dimensional conditional distributions of the i-th element of $\mathbf{w}$ conditioned on the remaining elements $\boldsymbol{\mu}_{\neg i,tr} = [\mu_{1,tr}, \ldots, \mu_{i-1,tr}, \mu_{i+1,tr}, \ldots, \mu_{N,tr}]^T$. These conditional distributions are expressed as, [4],

$$w_i|\boldsymbol{\mu}_{\neg i,tr} \sim \mathcal{N}_{\mathbf{R}_+^1}\left(w_i|\mu_i^*, \sigma_{ii}^{*2}\right) \tag{20}$$

with

$$\mu_i^* = \mu_i + \boldsymbol{\sigma}_{\neg i}^T \boldsymbol{\Sigma}_{\neg i \neg i}^{-1}\left(\boldsymbol{\mu}_{\neg i,tr} - \boldsymbol{\mu}_{\neg i}\right) \tag{21}$$

$$\sigma_{ii}^{*2} = \sigma_{ii} - \boldsymbol{\sigma}_{\neg i}^T \boldsymbol{\Sigma}_{\neg i \neg i}^{-1} \boldsymbol{\sigma}_{\neg i}, \tag{22}$$

where μ_i and σ_{ii} represent the i-th and ii-th elements of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively, the $(N-1)\times(N-1)$ matrix $\boldsymbol{\Sigma}_{\neg i \neg i}$ is formed by removing the i-th row and the i-th column from $\boldsymbol{\Sigma}$, while the $(N-1)\times 1$ vector $\boldsymbol{\sigma}_{\neg i}$ is the ith column of $\boldsymbol{\Sigma}$ after removing its ith element, and $\boldsymbol{\mu}_{\neg i}$ is the vector resulting from $\boldsymbol{\mu}$ after removing its i-th element μ_i. The j-th iteration of the proposed scheme can be expressed as

$$\begin{aligned}
&1.\ \mu_{1,tr}^{(j)} = \mathrm{E}[p(w_1|\mu_{2,tr}^{(j-1)}, \mu_{3,tr}^{(j-1)}, \ldots, \mu_{N,tr}^{(j-1)})] \\
&2.\ \mu_{2,tr}^{(j)} = \mathrm{E}[p(w_2|\mu_{1,tr}^{(j)}, \mu_{3,tr}^{(j-1)}, \ldots, \mu_{N,tr}^{(j-1)})] \\
&\qquad\vdots \\
&N.\ \mu_{N,tr}^{(j)} = \mathrm{E}[p(w_N|\mu_{1,tr}^{(j)}, \mu_{2,tr}^{(j)}, \ldots, \mu_{N-1,tr}^{(j)})].
\end{aligned} \tag{23}$$

Note that in the one-dimensional case, the expectation of a random variable $x \sim \mathcal{N}_{\mathbf{R}_+^1}(x|\mu^*, \sigma^{*2})$, such as those in (23), can be computed as, [4],

$$\mathrm{E}[x] = \mu^* + \frac{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{\mu^{*2}}{\sigma^{*2}}\right)}{1-\frac{1}{2}\mathrm{erfc}\left(\frac{\mu^*}{\sqrt{2}\sigma^*}\right)}\sigma^*, \tag{24}$$

with $\mathrm{erfc}(\cdot)$ being the complementary error function. It has been experimentally verified that this scheme converges after a few iterations, [4].

In the sequel we show that it is possible to drop the dependence on $\boldsymbol{\mu}$ in (21) and sidestep the complex operations of matrix inversions, i.e., the computation of $\boldsymbol{\Sigma}_{\neg i \neg i}^{-1} \forall i$, which have complexity $\mathcal{O}(N(N-1)^3)$. To this end, straightforward computations for μ_i^* in (21) yield that

$$\mu_i^* = \boldsymbol{\sigma}_{\neg i}^T \boldsymbol{\Sigma}_{\neg i \neg i}^{-1} \boldsymbol{\mu}_{\neg i,tr} + \begin{bmatrix} -\boldsymbol{\sigma}_{\neg i}^T \boldsymbol{\Sigma}_{\neg i \neg i}^{-1} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_{\neg i} \\ \mu_i \end{bmatrix}. \tag{25}$$

Setting $\mathbf{z} = \boldsymbol{\Phi}^T\mathbf{y}$, (10) becomes $\boldsymbol{\mu} = \langle\beta\rangle\boldsymbol{\Sigma}\mathbf{z}$ and we get

$$\begin{bmatrix} \boldsymbol{\mu}_{\neg i} \\ \mu_i \end{bmatrix} = \mathbf{T}_i\boldsymbol{\mu} = \langle\beta\rangle\mathbf{T}_i\boldsymbol{\Sigma}\mathbf{z} = \langle\beta\rangle\mathbf{T}_i\boldsymbol{\Sigma}\mathbf{T}_i^T\mathbf{T}_i\mathbf{z} = \langle\beta\rangle\boldsymbol{\Sigma}_i(\mathbf{T}_i\mathbf{z}) \tag{26}$$

where $\mathbf{T}_i$ is an appropriate permutation matrix and $\boldsymbol{\Sigma}_i$ is obtained from $\boldsymbol{\Sigma}$ by moving its i-th column and row to the end of the matrix,

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \boldsymbol{\Sigma}_{\neg i \neg i} & \boldsymbol{\sigma}_{\neg i} \\ \boldsymbol{\sigma}_{\neg i}^T & \sigma_{ii} \end{bmatrix}. \tag{27}$$

By substituting (27) in (26), and then in (25), it easily follows that

$$\mu_i^* = \boldsymbol{\sigma}_{\neg i}^T \boldsymbol{\Sigma}_{\neg i \neg i}^{-1} \boldsymbol{\mu}_{\neg i,tr} + \langle\beta\rangle\sigma_{ii}^{*2} z_i \tag{28}$$

Let us denote with $\mathbf{v}_{\neg i}^T$ the i-th row of $\mathbf{V} = \langle\beta\rangle^{-1}\boldsymbol{\Sigma}^{-1}$ excluding its i-th element v_{ii}. From (27) and the partitioned covariance matrix inversion formula, we get

$$\mathbf{v}_{\neg i}^T = -\frac{\langle\beta\rangle^{-1}}{\sigma_{ii}^*}\boldsymbol{\sigma}_{\neg i}^T\boldsymbol{\Sigma}_{\neg i \neg i}^{-1}, \qquad \sigma_{ii}^{*2} = \frac{\langle\beta\rangle^{-1}}{v_{ii}}. \tag{29}$$

Using (29), (28) becomes

$$\mu_i^* = \frac{1}{v_{ii}}(z_i - \mathbf{v}_{\neg i}^T\boldsymbol{\mu}_{\neg i,tr}), \tag{30}$$

that is, each μ_i^* is efficiently computed with N operations. The proposed algorithm is summarized in Table 1. Note that matrix inversions have been completely eliminated and the required computational complexity of the algorithm is $\mathcal{O}(N^2)$ per iteration t, which is one order of magnitude less than the original BI-ICE algorithm, [4]. In addition, both algorithms converge very fast, exhibit similar estimation performance, and produce sparse estimates without the need of tuning or cross-validating any parameters.

Input $\mathbf{y}, \boldsymbol{\Phi}$
Initialize $\beta, \boldsymbol{\gamma}, \boldsymbol{\lambda}$
Compute $\mathbf{A} = \boldsymbol{\Phi}^T\boldsymbol{\Phi}$, and $\mathbf{z} = \boldsymbol{\Phi}^T\mathbf{y}$
for $t = 1, 2, \ldots$
- compute $\mathbf{V}(t) = \mathbf{A} + \boldsymbol{\Gamma}^{-1}(t)$
for $i = 1, 2, \ldots, N$
- extract $\mathbf{v}_{\neg i}(t)$ and $v_{ii}(t)$ from $\mathbf{V}(t)$
- compute $\sigma_{ii}^{*2}(t)$ from (29) and $\mu_i^*(t)$ from (30)
- compute $\mu_{i,tr}(t)$ from (24)
end for
- compute $\beta(t)$ from (19)
- compute $\boldsymbol{\gamma}(t)$ from (15)
- compute $\boldsymbol{\lambda}(t)$ from (16)
end for

Table 1: The proposed fast variational Bayes algorithm.

5. EXPERIMENTAL RESULTS

In this section, we test the proposed VB algorithm on the calibrated OMEGA cube of the Syrtis Major area used in [5]. The endmember matrix $\mathbf{\Phi}$ contains the spectral signatures of 32 mineral, previously detailed in [5]. Syrtis Major is a Hesperian volcanic complex composed essentially of basalts. Mafic minerals, olivine and both low-calcium (LCP) and rich-calcium (HCP) pyroxenes [7], as well as phyllosilicates [8] have been identified in the area. Moreover, there is a significant presence of hydrated minerals [9], feldspar [8] and iron-bearing minerals such as iron oxides [10]. In the study area, Mustard et. al. in [7] have already identified the presence of three specific mafic minerals (i) hypersthene, (ii) diopside and (iii) fayalite.

Abundance maps obtained by applying the proposed algorithm reveal the presence of three areas with distinct characteristics in the image. In the middle part of the image, LCP pyroxenes prevail, as shown in Fig. 1a. The results shown in Fig. 1a are in accordance to the corresponding maps in [5]. The upper part of the image presents low reflectance and low abundance values and no spatially predominant mineral. HCP diopside is observed only in this area but in localized outcrops, as shown in Fig. 1b. In addition, the lower zone of the image is characterized by strong presence of iron oxides, such as magnetite (Fig. 1c) and hematite, accompanied by clay minerals such as nontronite. Finally, olivines are detected in few pixels with low abundance values in the middle left part of the image, while phyllosilicates such as muscovite are detected in the whole image, although having low abundances. The latter results are not shown here due to space limitations. Furthermore, the mean number of abundances of value higher than 0.1 is 1.74, i.e., in the mean, approximately two endemembers are present in each pixel, which justifies the use of a sparsity-promoting unmixing scheme. The sum of abundances per pixel in the upper half of the image varies around 0.35 while in the bottom half the same sum exceeds 1.5. This is possible since the sum-to-one constraint, i.e., $\sum_{i=1}^{N} w_i = 1$, is not imposed in the proposed model and is an indication that the endmembers library used in the unmixing process may be insufficient to effectively describe the exact mineral composition of the scene, as also noted in [5].

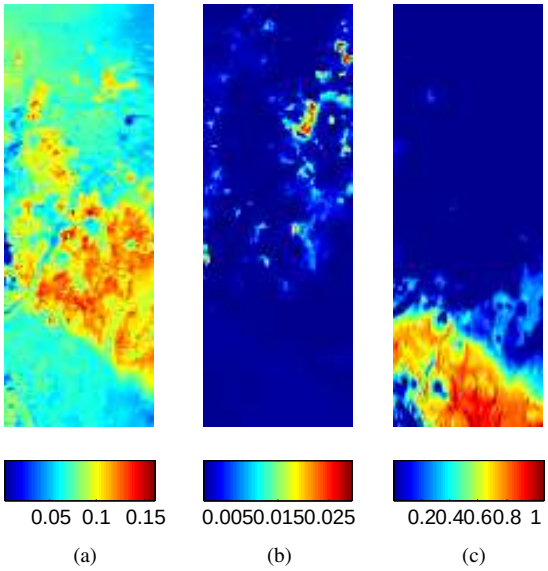

Fig. 1: Abundance maps of (a) hypersthene (b) diopside and (c) magnetite in Syrtis Major obtained using the proposed VB algorithm.

6. CONCLUSION

In this paper, an iterative VB algorithm for the sparse linear unmixing problem is proposed. The algorithm converges very fast and has a low computational complexity of $\mathcal{O}(N^2)$ per iteration. Experimental results conducted on real hyperspectral data captured from the OMEGA sensor clearly demonstrate that the proposed method favors sparse estimates and allows for the reliable mineral analysis and mapping of the martian surface.

7. ACKNOWLEDGMENTS

The authors would like to thank ESA and the OMEGA team for providing us the hyperspectral data used in this paper.

8. REFERENCES

[1] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Qian Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 2, pp. 354 –379, April 2012.

[2] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 6, pp. 2014 –2039, June 2011.

[3] M.-D. Iordache, J.M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 11, pp. 4484 –4502, Nov. 2012.

[4] K.E. Themelis, A.A. Rontogiannis, and K.D. Koutroumbas, "A novel hierarchical bayesian approach for sparse semisupervised hyperspectral unmixing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 2, pp. 585 –599, Feb. 2012.

[5] F. Schmidt et al., "Accuracy and performance of linear unmixing techniques for detecting minerals on omega/mars express," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, June 2011, pp. 1 –4.

[6] D.G. Tzikas, A.C. Likas, and N.P. Galatsanos, "The variational approximation for Bayesian inference," *Signal Processing Magazine, IEEE*, vol. 25, no. 6, pp. 131 –146, Nov. 2008.

[7] J.F. Mustard et al., "Olivine and pyroxene diversity in the crust of Mars," *Science*, vol. 307, pp. 1594–1597, 2005.

[8] F. Poulet et al., "Phyllosilicates on Mars and implications for early martian climate," *Nature*, vol. 438, no. 7068, pp. 623–627, 2005.

[9] J.P. Bibring et al., "Mars surface diversity as revealed by the OMEGA/Mars express observations," *Science*, vol. 307, pp. 1576–1581, 2005.

[10] J.F. Mustard et al., "The surface of Syrtis Major: Composition of the volcanic substrate and mixing with altered dust and soil," *Journal of Geophysical Research*, vol. 98, no. E2, pp. 3387–3400, 1993.